





OOoCon 2004 - Berlin



**Translation Technology At  
Sun Microsystems, Inc.  
BOF Session**




# Agenda

- Our interest in translation tools
  - Explanation of common problems facing translators
  - Some common standards
    - *TMX & XLIFF*
  - Our tools
  - Demo of our translation editor
- 
- 



# Our interest in translation tools

- Annual throughput of 40 million words
  - Translation into 9 languages (sometimes more)
  - Wide variety of products translated, from technical and developer material to end-user focused material
- 



# Problems facing translators

- New file format “arms race”
  - *Constantly encountering new file formats*
  - *Not all translators are technical and some file formats can be complex*
  - *Tools authors also need to maintain tool to keep up with changing formats*
- Reuse existing translations
- Use common terminology when translating
- Maintain consistency between documentation and software translations
- Avoid breaking files
  - *Formatting problems*
  - *Encoding problems*
  - *Missing messages*



# Common translation standards

- TMX (Translation Memory eXchange)
  - *Simple XML based container format*
  - *Provides basic source language / target language mapping & some context information*



# Common translation standards

- XLIFF (eXtensible Localization Interchange File Format)
  - *XML based container format*
  - *Provides more complex features than TMX :*
    - *De-coupling translatable text from formatting*
    - *Alternative translations from TM/MT/Glossary tools*
    - *Additional context information*
    - *Comments from translators*
    - *Format highlighting*



# Our tools

- Translation memory tool
  - *J2EE application including filters to convert incoming formats to XLIFF*
  - *Fast exact & fuzzy lookup against a very large database of pretranslated text (1.4 million source segments, ~ 8 million translated segments)*
  - *Statistics output file*
- Translation editor
  - *Edits XLIFF files*
  - *Format highlighting*
  - *Repetition & local “mini-TM” functionality*
  - *Spell checkers & format checking*



# Demo & Resources

- TMX [1]
- XLIFF [2]
- GADC [3]
- Tim's Web Log :- ) [4]

[1] <http://www.lisa.org/tmx>

[2] [http://www.oasis-open.org/committees/  
/tc\\_home.php?wg\\_abbrev=xliff](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xliff)

[3] [http://developers.sun.com/techttopics/  
/global](http://developers.sun.com/techttopics/global)

[4] <http://blogs.sun.com/timf>